Stochastic Processes and their Applications 34 (1990) 39–47
North-Holland

A MULTIPLICATIVE ERGODIC THEOREM FOR LIPSCHITZ MAPS

John H. ELTON

School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, USA

Received 8 December 1988

If $(F_n, n \geqslant 0)$ is a stationary (ergodic) sequence of Lipschitz maps of a locally compact Polish space X into itself having a.s. negative Lyapunov exponent function, the compostition process $F_n \cdots F_1 x$ converges in distribution to a stationary (ergodic) process in X (independent of x). For every x, the empirical distribution of a trajectory converges with probability one, and for every $\varepsilon > 0$, almost every trajectory is eventually within ε of the support. We use the fact that the Lyapunov exponent of a process "run backwards" is the same as forwards. A set invariance condition is given for the case when (F_n) is a Markov chain. The result has applications to computer graphics and stability in control theory.

AMS 1980 Subject Classifications: Primary 28D99, 60G10, 60G17; Secondary 58F11, 60J05, 60F05, 60F15.

stationary sequence * Lipschitz map * Lyapunov exponent * ergodic theorem

1. Stationary processes

Let (X, d) be a complete, separable locally compact metric space. Let $C(X, X)$ be the continuous maps from X into X, and let $\mathrm{Lip}(X, X)$ be the Lipschitz maps from X into X. We are interested in multiplying (i.e., composing) maps chosen from $\mathrm{Lip}(X, X)$ according to a stationary distribution.

Now $C(X, X)$ with the compact-open topology (that is, topology of uniform convergence on compact sets) is a complete metric space (Husain, 1977, Ch. VIII), and is also separable—this is shown in Kuratowski (1952, p. 120) for compact X, and easily extends to our case by the σ-compactness of X. For linear maps from $\mathbb{R}^n \to \mathbb{R}^n$ this is equivalent to the usual norm topology.

For $f \in \mathrm{Lip}(X, X)$, define

$$\|f\| = \sup_{x \neq y}(d(fx, fy)/d(x, y)).$$

This is Borel measurable since X is separable ($\mathrm{Lip}(X, X)$ inherits its topology from $C(X, X)$ and will have the Borel σ-algebra). For affine maps this is the norm of the linear part.

Research supported in part by DARPA "Image Compression Project," ONR Contract #N00014-86-C-0446.

We shall use the idea of extending a stationary process into the past, so that it may be considered to have been going on forever. The next lemma appears in Doob (1953, pp. 456–458), stated for real-valued processes. That proof, which uses Kolmogorov's extension theorem, is valid here since Kolmogorov's theorem applies in the setting of a complete separable metric space. Recall that a process $(Y_n, n \geq 0)$ (resp. $(Y_n, -\infty < n < \infty)$) with values in M is *stationary* (ergodic) iff the left shift on $M^{\mathbb{N}}$ (resp. $M^{\mathbb{Z}}$) is measure-preserving (ergodic) for the measure $P \circ Y^{-1}$, where $(\Omega, \mathscr{F}, P)$ is the underlying probability space.

Lemma 1. *Let $(Y_n, n \geq 0)$ be a stationary process in a complete separable metric space. Then there exists a stationary process $(\tilde{Y}_n, -\infty < n < \infty)$ such that $(\tilde{Y}_n, n \geq 0)$ has the same distribution as $(Y_n, n \geq 0)$. Also, $(\tilde{Y}_n, -\infty < n < \infty)$ is ergodic iff $(Y_n, n \geq 0)$ is ergodic.* □

The next proposition is the Furstenberg–Kesten theorem (Furstenberg and Kesten, 1960) in the setting of Lipschitz maps rather than linear maps; we also show that the Lyapunov exponent function is the same if the process is run backwards.

Proposition 2. (i) *Let $(F_n, n \geq 0)$ be a stationary process in* $\mathrm{Lip}(X, X)$ *such that $\mathbb{E} \log^+ \|F_0\| < \infty$. Then there exists an invariant function $\chi : \Omega \to \mathbb{R} \cup \{-\infty\}$ (called the Lyapunov exponent) with $\chi^+ \in L_1(P)$ such that*

$$\frac{1}{n} \log\|F_n \cdots F_1\| \to \chi \quad a.s.$$

and

$$\lim_{n\to\infty} \frac{1}{n} \int \log\|F_n \cdots F_1\| \, \mathrm{d}P = \inf \frac{1}{n} \int \log\|F_n \cdots F_1\| \, \mathrm{d}P = \int \chi \, \mathrm{d}P.$$

(ii) *Furthermore, if $(F_n, -\infty < n < \infty)$ extends $(F_n, n \geq 0)$ into the past, then for all k,*

$$\lim_{n\to\infty} \frac{1}{n} \log\|F_{k-1} \cdots F_{k-n}\| = \lim_{n\to\infty} \frac{1}{n} \log\|F_n \cdots F_1\| = \chi \quad a.s.$$

Proof. (i) Follows from Kingman's subadditive ergodic theorem exactly as in the case of linear maps, as presented in Krengel (1985, p. 40).

To prove (ii), we may consider w.l.o.g.,

$$F_n(\omega) = F(\tau^n \omega)$$

where τ is an invertible measure-preserving transformation on $(\Omega, \mathscr{F}, P)$ and $F : \Omega \to \mathrm{Lip}(X, X)$ is measurable. Then the same proof as in (i), using Kingman's theorem and *writing down the composition in reverse order*, shows that

$$\frac{1}{n} \log\|F_{k-1}(\omega) \cdots F_{k-n}(\omega)\| = \frac{1}{n} \log\|F(\tau^{k-1}\omega) \cdots F(\tau^{k-n}\omega)\| \to \tilde{\chi}(\omega),$$

say, a.s., where $\tilde{\chi}$ is a τ-invariant function. Assume L_1 convergence as well. Then

$$\int \left| \frac{1}{n} \log \| F(\tau^{k-1}\omega) \cdots F(\tau^{k-n}\omega) \| - \tilde{\chi}(\omega) \right| \mathrm{d}P(\omega)$$

$$= \int \left| \frac{1}{n} \log \| F(\tau^{n}\omega) \cdots F(\tau\omega) \| - \tilde{\chi}(\omega) \right| \mathrm{d}P(\omega),$$

since τ is measure-preserving and $\tilde{\chi}$ is invariant. By the uniqueness of L_1 limits, it follows that $\tilde{\chi} = \chi$ a.s. Finally, a truncation argument can be used to reduce the general case to our assumption of L^1 convergence. □

Remark. Note that if the process in Proposition 2 is *ergodic* then χ is a.s. constant, and then $\chi < 0$ iff for some n, $\int \log \| F_n \cdots F_1 \| \, \mathrm{d}P > 0$. This is often a practical way of determining if $\chi < 0$.

We now state our main result (Theorem 3(iv) below). The proof is made very simple by the right use of the time-reversal idea. Similar results were obtained in Barnsley and Elton (1989) for i.i.d. sequences of maps with distribution supported on *finitely* many maps, which was generalized in Berger and Amit (1989) to i.i.d. sequences of affine maps (in Barnsley and Elton, 1989, the negative Lyapunov condition was only given in the form in the remark above). See also Barnsley, Elton and Hardin (1989) for the case of a finite Markov chain of maps. All three proofs above used a version of the time-reversal idea. Our proof of (i) below generalizes the proof in Berger and Amit (1989). We feel that the stationary setting is the "right" one, and gives a better understanding of the meaning of the time-reversal. We also feel this proof clarifies, in a general setting, the connection between the type of iteration $F_1 F_2 \cdots F_n x$ used in symbolic dynamics and the natural stochastic type of iteration $F_n F_{n-1} \cdots F_1 x$. Note that an i.i.d. sequence of maps does not give rise to an independent process in X (it is Markov), and a stationary Markov sequence of maps does not give rise to a Markov process in X; but a stationary process of maps becomes, under composition with the right starting random variable in X, a stationary process in X.

Theorem 3. *Let* $(F_n, n \geqslant 0)$ *be a stationary process in* $\mathrm{Lip}(X, X)$ *satisfying*

$$\int \log^+ \| F_0 \| \, \mathrm{d}P < \infty \quad \textit{and} \quad \int \log^+ d(F_0 x_0, x_0) \, \mathrm{d}P < \infty$$

for some x_0.

Suppose $\chi < 0$ *a.s. W.l.o.g. assume* (F_n) *extends backward in time to* $(F_n, -\infty < n < \infty)$ *as in Lemma* 1. *Then*

(i) *For any* $x \in X$, $-\infty < k < \infty$,

$$Y_k = \lim_{n \to \infty} F_k F_{k-1} \cdots F_{-n} x$$

exists a.s. and is independent of x, and is a stattionary process in X. Note that $Y_k = F_k F_{k-1} \cdots F_1 Y_0$, $k \geq 0$, *so* Y_0 *is a random variable of starting values that makes the multiplicative process stationary.*

(ii) *For any* $x \in X$,

$$(F_n F_{n-1} \cdots F_1 x, F_{n+1} F_n \cdots F_1 x, F_{n+2} F_{n+1} \cdots F_1 x, \dots)$$

converges in distribution to $(Y_0, Y_1, \dots)$. *In particular*

$$F_n F_{n-1} \cdots F_1 x \xrightarrow{\mathcal{D}} Y_0.$$

(iii) *If* (F_n) *is ergodic, then so is* (Y_n). *In fact,* (Y_n) *is a factor of* (F_n), *in the sense of ergodic theory.*

(iv) (*Trajectories.*) *For every x, for a.a.* ω,

$$\frac{1}{n} \sum_{k=1}^{n} f(F_k(\omega) \cdots F_1(\omega) x) \to \mathbb{E}(f(Y_0) \mid \mathcal{I})(\omega)$$

for all bounded continuous real-valued functions on X, where $\mathcal{I}$ *is the* σ*-algebra of invariant events for* (F_n).

In particular, if (F_n) *is ergodic, then the empirical distribution of the trajectories of the multiplicative process starting at any x converge weakly to the measure* $\mu = P \circ Y_0^{-1}$, *a.s.*

(v) (*Convergence of trajectories to support.*) *Let A be the support of* $\mu = P \circ Y_0^{-1}$. *Let* $x \in X$ *and let* $\varepsilon > 0$. *Then for a.a.* ω, *there exists* $n_0(\omega)$ *such that* $n \geq n_0 \Rightarrow d(F_n \cdots F_1 x, A) < \varepsilon$. *Thus in the ergodic case, A can be characterized as follows:* $y \in A$ *iff for every* $x \in X$, *for every neighborhood G of y,* $F_n \cdots F_1 x$ *visits G infinitely often, a.s.*

Remarks. From (iii) it follows that if (F_n) is i.i.d., then even though (Y_n) is not independent, it *is* isomorphic to a Bernoulli shift. Similarly, if (F_n) is a Markov shift, then (Y_n) is isomorphic to a Markov shift, even though it is not Markovian.

(v) is stronger than would be expected from (iv). (iv) implies that if $y \notin A$, there is a neighborhood of y such that the proportion of time a trajectory spends in the neighborhood approaches 0, for almost all trajectories. (v) makes the stronger assertion that almost all trajectories visit it only *finitely* many times. This stronger result follows from the time reversal argument.

If x were chosen according to the measure μ, (iv) would be just a statement of the classical pointwise ergodic theorem. The significance of (iv) is that one does not have to know μ in advance to pick a starting value to generate a trajectory that will "draw a picture" of μ.

Proof of Theorem 3. It is easy to show by the triangle inequality that the hypotheses imply $\int \log^+ d(F_0 x, x)\, \mathrm{d}P < \infty$ for all x. Now

$$d(F_k F_{k-1} \cdots F_{-n} x, F_k F_{k-1} \cdots F_{-n-1} x) \leq \|F_k \cdots F_{-n}\|\, d(x, F_{-n-1} x).$$

For each j, let $\Omega_j = \{\chi < -1/j\}$, so $P(\bigcup \Omega_j) = 1$. By Proposition 2(ii), for each $\omega \in \Omega_j$, $(1/n)\log\|F_k(\omega)\cdots F_{-n}(\omega)\| < -1/j$ for sufficiently large n (depending on ω); i.e., $\|F_k(\omega)\cdots F_{-n}(\omega)\| < \exp(-n/j)$. Since $\mathbb{E}\log^+ d(F_0x, x) < \infty$, it follows that

$$\sum_{n=1}^{\infty} P(\log^+ d(F_{-n-1}x, x) > n/2j) < \infty,$$

so for a.a. $\omega \in \Omega_j$, $d(F_{-n-1}x, x) \leq \exp(n/2j)$ eventually. Thus for a.a. $\omega \in \Omega_j$,

$$d(F_kF_{k-1}\cdots F_{-n}x, F_k\cdots F_{-n-1}x) < \exp(-n/2j)$$

for sufficiently large n, so $F_kF_{k-1}\cdots F_{-n}x$ is a Cauchy sequence and converges, to say Y_k. Also, $d(F_k\cdots F_{-n}x, F_k\cdots F_{-n}x') \leq \|F_k\cdots F_{-n}\| d(x, x') \to 0$, so Y_k is independent of x. Since $P(\bigcup \Omega_j) = 1$, this proves the first statement in (i).

The stationarity of (Y_k) follows easily from that of (F_n).

To prove (ii), note that for each n, $(F_n\cdots F_1x, F_{n+1}\cdots F_1x, \ldots)$ has the same distribution as $(F_0F_{-1}\cdots F_{-n+1}x, F_1F_0\cdots F_{-n+1}x, \ldots)$ by stationarity, and the latter converges to $(Y_0, Y_1, \ldots)$ by (i).

(iii) is most easily seen from the measure-preserving transformation point of view. If (F_n) is ergodic, we may consider $F_n(\omega) = F(\tau^n\omega)$ where τ is an ergodic invertible m.p.t. and $F: \Omega \to \mathrm{Lip}(X, X)$ is measurable. Then

$$Y_k(\omega) = \lim_{n\to\infty} F(\tau^k\omega)\cdots F(\tau^{-n}\omega)x,$$

so $Y_0(\tau^k\omega) = Y_k(\omega)$. It now follows (Doob, 1953, p. 457) that (Y_k) is an ergodic process.

To prove (iv), first let f be a real-valued continuous function on X with compact support. Then

$$\frac{1}{n}\sum_{k=1}^{n} f(F_k\cdots F_1Y_0) \to \mathbb{E}(f(Y_0)\,|\,\mathcal{I}) \quad \text{a.s.}$$

by the pointwise ergodic theorem. Also,

$$d(F_k\cdots F_1Y_0, F_k\cdots F_1x) \leq \|F_k\cdots F_1\|\, d(Y_0, x) \to 0 \quad \text{a.s.}$$

since $\chi < 0$ a.s. From this it is clear by the uniform continuity of f that

$$\frac{1}{n}\sum_{k=1}^{n} f(F_k\cdots F_1x) \to \mathbb{E}(f(Y_0)\,|\,\mathcal{I}) \quad \text{a.s.}$$

Since the real continuous functions with compact support are separable, it follows by a 3ε argument that almost all ω are "good" for all such f simultaneously. Then the result for bounded real continuous f follows by a standard argument using Urysohn's lemma (note $(F_k\cdots F_1x)$ is tight since it converges in distribution).

To prove (v) choose any $y \in A$. For a.a. ω, there exists $n_0(\omega)$ such that $n \geq n_0 \Rightarrow \|F_n\cdots F_1\| < \varepsilon/(3(1 + d(x, y)))$. For a.a. ω, $Y_n = \lim_{k\to\infty} F_n\cdots F_{-k}y \in A$ for all n; and $\{\lim_{k\to\infty} F_0(\omega)\cdots F_{-k}(\omega)y\}$ is dense in A for any set of ω of probability 1. Thus for a.a. ω, we can find $n_0(\omega)$ such that: $n \geq n_0 \Rightarrow \|F_n\cdots F_1\| < \varepsilon/(3(1 + d(x, y)))$

and for each $n \geq n_0$ there exists k such that $d(F_n \cdots F_{-k}y, A) < \frac{1}{3}\varepsilon$ and $d(F_0 \cdots F_{-k}y, y) < 1$. Thus

$$\begin{aligned} d(F_n \cdots F_1 x, A) &\leq d(F_n \cdots F_1 x, F_n \cdots F_1 y) + d(F_n \cdots F_1 y, F_n \cdots F_{-k} y) \\ &\quad + d(F_n \cdots F_{-k} y, A) \\ &< \tfrac{1}{3}\varepsilon + \tfrac{1}{3}\varepsilon d(F_0 \cdots F_{-k} y, y) + \tfrac{1}{3}\varepsilon < \varepsilon. \qquad \square \end{aligned}$$

2. Markov processes

If (F_n) is a stationary *Markov* process in $\text{Lip}(X, X)$, then (Y_n) is not in general Markov, but $((Y_n, F_n))$ is a stationary Markov process in $X \times \text{Lip}(X, X)$, with transition probability function $\tilde{p}((x, F), C)$ = probability of transfer from (x, F) into C in one step $= \int I_C(Gx, G)p(F, \mathrm{d}G)$ where p is the transition probability function for the Markov process (F_n).

Now suppose F_0 has finite support $\{\gamma_1, \ldots, \gamma_M\}$ so that (F_n) is a stationary Markov chain with finite state space. Of special interest is the case when (F_n) is an irreducible chain and some maps are not allowed to follow others, so that there are zeroes in the transition probability matrix (p_{ij}) for the chain. Then these processes make contact with "subshifts of finite type" in symbolic dynamics, as was discussed in Barnsley, Elton and Hardin (1989). If X is compact and $\{\gamma_1, \ldots, \gamma_M\}$ are uniform contractions, then $Y(\underline{i}) = \lim_{n\to\infty} \gamma_{i_1} \cdots \gamma_{i_n} x$ exists for *every* code sequence $\underline{i} = (i_1, i_2, \ldots)$, $1 \leq i_j \leq M$, and is independent of x. The values of $Y(\underline{i})$ as $\underline{i}$ ranges over *allowable* code sequences—that is, ones for which $p_{i_j i_{j+1}} > 0$ for all j—correspond to the support of Y_0 in Theorem 3 above; by the *support* of Y_0 we mean the closed set

$$S = \{y \in X \colon P(d(Y_0, y) < \varepsilon) > 0 \text{ for all } \varepsilon > 0\}.$$

Then S can be written as the union of compact sets $\{S_j\}_{j=1}^M$ which satisfy the following interesting invariance relation, as shown in Barnsley, Elton and Hardin (1989):

$$S_j = \bigcup_{i:\, p_{ij} > 0} \gamma_j(S_i). \tag{$*$}$$

Results on the fractal dimension of S are given in Barnsley, Elton and Hardin (1989). Furthermore, $\{S_j\}$ are the unique non-empty compact sets which satisfy $(*)$. This generalized the simple invariance relation $S = \bigcup_{j=1}^M \gamma_j(S)$ which holds for the support in the i.i.d. case, see Hutchinson (1981). S is called the *attractor* in those contexts. It was shown in Berger and Amit (1989) for the i.i.d. case with affine maps that when expansive maps are allowed, the support S still satisfies a similar invariance relation, namely, S is the minimum closed non-empty set satisfying $\bigcup_{j=1}^M \gamma_j(S) \subset S$.

We would like to show that a similar invariance relation to $(*)$ holds, when expansive maps are allowed, for an irreducible Markov chain (note this is an *ergodic* process).

Theorem 4. *Let $(F_n, n \geq 0)$ be an irreducible stationary Markov chain on a finite set of maps $\{\gamma_1, \ldots, \gamma_M\} \subset \text{Lip}(X, X)$ such that the Lyapunov exponent $\chi < 0$. Let S be the support of Y_0 in Theorem* 3.

Then there exist non-empty closed sets $S_1, \ldots, S_M$ such that $S = \bigcup_{j=1}^{M} S_j$ and

$$\bigcup_{i: p_{ij}>0} \gamma_j(S_i) \subset S_j \subset \bigcup_{i: p_{ij}>0} \overline{\gamma_j(S_i)} \tag{$**$}$$

for $j = 1, \ldots, M$. Furthermore, if $S_1', \ldots, S_M'$ are non-empty closed sets satisfying $\bigcup_{i: p_{ij}>0} \gamma_j(S_i') \subset S_j'$ for all j, then $S_j \subset S_j'$ for all j, so the S_j are minimum sets.

If the γ_j are closed maps, such as non-singular affine maps, then we get $\bigcup_{i: p_{ij}>0} \gamma_j(S_i) = S_j$ for all j.

Proof. Let

$$S_j = \{y \in X\text{: for all } \varepsilon > 0,\ P(d(Y_1, y) < \varepsilon \text{ and } F_1 = \gamma_j) > 0\}.$$

S_j is obviously closed. Let $y \in S$. For each n, $P(d(Y_1, y) < 1/n$ and $F_1 = \gamma_{j_n}) > 0$ for some j_n since F_1 must take on one of the values $\gamma_1, \ldots, \gamma_M$. For some subsequence $j_{n_k} = j$ for all k, so it follows that $y \in S_j$, and thus $S = \bigcup_{j=1}^{M} S_j$. Since the chain is irreducible $P(F_1 = \gamma_j) > 0$ for all j, so it follows by the Lindelöf property of X that $S_j \neq \emptyset$.

Now let $y \in S_i$ and assume $p_{ij} > 0$. For all $\varepsilon > 0$, $P(d(Y_0, y) < \varepsilon$ and $F_0 = \gamma_i) > 0$ since (F_0, Y_0) has the same distribution as (F_1, Y_1). Thus

$$P(d(\gamma_i Y_{-1}, Y) < \varepsilon \text{ and } F_0 = \gamma_i) > 0 \quad \text{for all } \varepsilon > 0$$

(this is the same event). Now

$$\begin{aligned} &P(F_1 = \gamma_j,\ F_0 = \gamma_i,\ d(y, \gamma_i Y_{-1}) < \varepsilon) \\ &\quad = p_{ij} P(F_0 = \gamma_i,\ d(y, \gamma_i Y_{-1}) < \varepsilon) \quad \text{(by the Markov property)} \\ &\quad > 0. \end{aligned}$$

But $F_1 = \gamma_j$ and $F_0 = \gamma_i \Rightarrow Y_1 = \gamma_j \gamma_i Y_{-1}$, so

$$P(d(Y_1, \gamma_j y) < \|\gamma_j\| \varepsilon \text{ and } F_1 = \gamma_j) > 0 \quad \text{for all } \varepsilon > 0,$$

soi $\gamma_j y \in S_j$. Thus $\bigcup_{i: p_{ij}>0} \gamma_j(S_i) \subset S_j$ as desired.

Going the other way, let $y \in S_j$. By the same argument as before, since $\{\gamma_1, \ldots, \gamma_M\}$ is finite, there is some $1 \leqslant i \leqslant M$ such that for all $\varepsilon > 0$,

$$P(d(Y_1, y) < \varepsilon \text{ and } F_1 = \gamma_j \text{ and } F_0 = \gamma_i) > 0.$$

Thus

$$p_{ij} > 0 \quad \text{and} \quad P(d(\gamma_j Y_0, y) < \varepsilon,\ F_0 = \gamma_i) > 0 \quad \text{for all } \varepsilon > 0.$$

By the Lindelöf property of X and the fact that (F_0, Y_0) has the same distribution as (F_1, Y_1), it follows that $P(F_0 = \gamma_i$ and $Y_0 \notin S_i) = 0$.Thus for all $\varepsilon > 0$, for some ω, $Y_0(\omega) \in S_i$ and $d(\gamma_j Y_0(\omega), y) < \varepsilon$. It follows that $y \in \overline{\gamma_j(S_i)}$, and so $(**)$ is proved.

To show minimality, suppose S'_j are non-empty closed sets satisfying $\bigcup_{i:p_{ij}>0} \gamma_j(S'_i) \subset S'_j$.

First, it is easy to see that there is $\Omega' \subset \Omega$ such that $P(\Omega') = 1$ and if $\omega_0 \in \Omega'$,

$$P(\omega\colon F_i(\omega) = F_i(\omega_0) \text{ and } F_{i-1}(\omega) = F_{i-1}(\omega_0)) > 0 \quad \text{for all } i.$$

Since as already observed $P(F_0 = \gamma_j \text{ and } Y_0 \notin S_j) = 0$, if $y \in S_j$ then for all $\varepsilon > 0$,

$$P(\omega\colon d(Y_0(\omega), y) < \varepsilon,\ F_0(\omega) = \gamma_j,\ Y_0(\omega) \in S_j,\ \omega \in \Omega') > 0.$$

So let $y = Y_0(\omega)$ where $\omega \in \Omega'$ and $y \in S_j$ and $F_0(\omega) = \gamma_j$; by what was just said, such y are dense in S_j. Now

$$y = \lim_{n\to\infty} F_0(\omega) F_{-1}(\omega) \cdots F_{-n}(\omega) x \quad \text{for any } x.$$

There is j_0 and a subsequence n_k such that $F_{-n_k}(\omega) = \gamma_{j_0}$ for all k. Since the chain is irreducible, $p_{i_0 j_0} > 0$ for some i_0. Let x be any element of S'_{i_0} (it is not empty by assumption).

Let $F_{-n}(\omega) = \gamma_{l_n}$ for all n. By definition of Ω', $p_{l_{n+1} l_n} > 0$ for all n. Applying the assumed invariance relation iteratively, starting with $n = n_k$, we get

$$F_{-n}(\omega) x = \gamma_{j_0} x \in S'_{j_0} = S'_{l_n},$$

then

$$F_{-n+1}(\omega) F_{-n}(\omega) x \in \gamma_{l_{n-1}}(S'_{l_n}) \subset S'_{l_{n-1}}, \quad \text{etc.},$$

arriving finally at

$$y = F_0(\omega) \cdots F_{-n}(\omega) x \in \gamma_j(S'_{l_1}) \subset S'_j.$$

Since such y are dense in S_j and S'_j is closed, we get $S_j \subset S'_j$ as desired. □

Remarks. The geometry of S seems to be rather restricted in the i.i.d. case supported on finitely many affine maps which includes at least one expansive map. It was shown in Barnsley and Elton (1989) for the i.i.d. case supported on finitely many 1-1 affine maps on $\mathbb{R}^1$ that if there is an expansive map among them, S is either a half-infinite interval or all of $\mathbb{R}^1$ (or a point). We have not proved it, but a similar filling in along rays seems to hold in $\mathbb{R}^n$, in non-degenerate cases.

However, for i.i.d. affine maps the measure itself may be strange even though the support is not. By studying the Fourier transform, one finds that the measure may be singular with respect to Lebesgue measure even when the support is an infinite interval in $\mathbb{R}^1$; for an example, see Barnsley and Elton (1989). This depends on the number-theoretical properties of the parameters of the affine maps, analogous to the situation discussed in Erdos (1939) and Garsia (1962) for the case of two contractive maps, involving P.V. numbers. Our work on this is incomplete and we will not go into it here.

3. Applications

Applications to computer graphics in the Markov chain case are discussed in Barnsley, Elton and Hardin (1989).

The result may be applied to give a generalization of the notion of stability in a common situation occuring in control theory. If the maps in Theorem 3 or 4 are matrices (linear maps on $\mathbb{R}^n$) which can be "controlled" to have negative Lyapunov exponent, the trajectories converge a.s. to zero. Our result shows that even when the maps have an additive term (affine maps), the trajectories will at least settle down to an attractor (and thus spend most of their time in a bounded set). Generalizations of this to continuous time (stochastic differential equations) should be of interest.

Acknowledgments

I would like to thank Michael Barnsley and Marc Berger for helpful discussions. Barnsley strongly advocated investigating the "running time backwards" point of view. I also thank Ullrich Krengel for helpful suggestions.

References

M. F. Barnsley and J. Elton, A new class of Markov processes for image encoding, Adv. in Appl. Probab. 20 (1988) 14-32.

M. F. Barnsley, J. Elton and D. Hardin, Recurrent iterated function systems, Constr. Approx. 5 (1989) 3-31.

M. Berger and Y. Amit, Products of random affine maps (1989), to appear.

J. L. Doob, Stochastic Process (Wiley, New York, 1953).

P. Erdos, On a family of symmetric Bernoulli convolutions, Amer. J. Math. 61 (1939) 974-976.

H. Furstenberg and H. Kesten, Products of random matrices, Ann. Math. Stat. 31 (1960) 457-469.

A. Garsia, Arithmetic properties of Bernoulli convolutions, Trans. AMS 102 (1962) 409-432.

T. Husain, Topology and Maps (Plenum, New York, 1977).

J. Hutchinson, Fractals and self-similarity, Indiana Univ. Math. J. 30 (1981) 713-747.

U. Krengel, Ergodic Theorems (de Gruyter, Berlin, 1985).

C. Kuratowski, Topologie, Vol. 1 (Polskie Towarzystwo Matematyczne, 1952).